AD-A113 663 MARYLAND UNIV COLLEGE PARK LAB FOR NUMERICAL ANALYSIS F/G 12/1
GENERALIZED FINITE ELEMENT METHODS: THEIR PERFORMANCE AND THEIR--ETC(U)
NOV 81 I BABUSKA, J E OSBORN N00014-77-C-0623
UNCLASSIFIED BN-975 NL

1 OF 1
AD
A113663

END
DATE
FILMED
5-82
DTIC

INSTITUTE FOR PHYSICAL SCIENCE AND TECHNOLOGY

Laboratory for Numerical Analysis

Technical Note BN-975

GENERALIZED FINITE ELEMENT METHODS: THEIR PERFORMANCE AND THEIR RELATION TO MIXED METHODS

by

I. Babuska

and

J. E. Osborn

November 1981

UNIVERSITY OF MARYLAND

(1)

Laboratory for Numerical Analysis

Technical Note BN-975

GENERALIZED FINITE ELEMENT METHODS: THEIR PERFORMANCE AND THEIR RELATION TO MIXED METHODS

by

I. Babuska

and

J. E. Osborn

November 1981

GENERALIZED FINITE ELEMENT METHODS: THEIR PERFORMANCE AND THEIR RELATION TO MIXED METHODS

by

I. Babuška

and

J. E. Osborn

Institute of Physical Science and Technology
Department of Mathematics
University of Maryland
College Park, MD 20742

The work of the first author was partially supported by ONR Contract N00014-77-C-0623 and that of the second by NSF Grant MCS-78-02851.

ABSTRACT

The notion of a generalized finite element method is introduced. This class of methods is analyzed and their relation to mixed methods is discussed. The class of generalized finite element methods offers a wide variety of computational procedures from which particular procedures can be selected for particular problems. A particular generalized finite element method which is very effective for problems with rough coefficients is discussed in detail.

1. Introduction

Although the general ideas involved in finite element methods (FEMs) are very old and are in fact closely related to classical variational principles (see [1], e.g., for a history of the finite element method), the full power of these methods is only seen in their implementation as computer codes. (For a list of over 650 codes for finite element analysis of structural mechanics problems see [2]). The success of finite element methods is due to several factors.

a) They allow for effective implementational procedures based on a collection process which leads to the construction of a sparse discretized system (stiffness matrix).

b) Finite element methods are robust. In many contexts this robustness is a consequence of underlying variational principles in mechanics.

c) These methods allow for the possibility of effective post-processing, as, e.g., in the calculation of stresses at arbitrary points.

d) Many finite element methods have a physical interpretation which permits an intuitive understanding of the computational process.

From the point of view of present day computational studies, properties a), b), and c) (especially a) and b)) are essential. Although d) is important, its importance has diminished because of progress in the mathematical understanding of finite element methods, the expanding field of applications for these methods,

and the progress in computer technology.

Since FEM s are increasingly used by unsophisticated users in a widening variety of fields, it is important to generalize the notion of the finite element method as far as possible while retaining the basic implementational architecture of the standard finite element methods (see a) above), so that a particular version can be selected for a class of problems so as to be maximally robust and to allow, if possible, for effective postprocessing.

This paper addresses the problem of generalizing the notion of the finite element method in the context of a simple one dimensional problem. We restrict ourselves to this model problem in order to clearly illustrate the ideas; the extension to two dimensional problems will be presented in a forthcoming paper. We thus consider the boundary value problem

$$\begin{cases} -(a(x)u')' + b(x)u = f, 0 < x < 1 \\ u(0) = u(1) = 0. \end{cases} \qquad (' = \frac{d}{dx})$$

Robustness of a FEM for this problem will mean that the method performs well independently of the smoothness of the coefficients a and b and the source term f. More precisely, we will say a method is robust with respect to a, b, and f if the rate of convergence is essentially governed by the approximation properties of the trial spaces we employ and the smoothness of the solution.

We will say that a method allows for postprocessing if it allows a significant improvement in accuracy to be obtained by a local procedure involving only a single element.

The concept of a generalized finite element method must, of course, include practically all known FEMs satisfying a), especially the standard FEM mixed methods, methods using different text and trial functions, and methods in which the shape of the trial functions is governed by the differential equations, as e.g., in methods for the convection-diffusion equation (see, e.g., [3] and references therein).

Section 2 introduces notation and preliminary notions. Section 3 introduces and discusses the Generalized Finite Element Method (GFEM). Section 4 analyses one special type of GFEM which is closely related to finite element methods based on L-splines (L here referring to a differential operator). This method is very appealing theoretically and is extremely robust, but in practice has various drawbacks. Nevertheless it provides valuable insight into what we could expect at "best." Section 5 analyses a modification of the method of Section 4 which better satisfies requirements of implementation. Section 6 analyses a further modification which is still very robust and is easily implementable. In Section 8 we show it is identical (in the sense of leading to the same approximate solution) to a certain mixed method. In addition, the method is shown to be very receptive to postprocessing. Section 7 discusses briefly the standard FEM (as a GFEM) and shows that it is unsuited for problems with rough coefficients. Also, the standard FEM is shown to be identical to another mixed method. Thus it is seen that various mixed methods (or versions of the mixed method) have substantially different robustness properties. In Section 9 we further elaborate the

method introduced in Section 6 and show that it "reacts" very well when the roughness of the coefficients is reduced. More precisely, we show that changing from measurable coefficients to coefficients with bounded variation improves the rate of convergence—as we should expect from a robust method. Section 10 contains some illustrative computations and Section 11 presents a summary of the conclusions of the paper.

Some of the results of this paper were announced in [4].

2. Preliminaries

Throughout this paper we will deal with one dimensional boundary value problems. Thus we consider the basic model Dirichlet problem

$$(2.1)\qquad \begin{cases} Lu \equiv -(au')' + bu = f, & 0 < x < 1 \\ u(0) = u(1) = 0. \end{cases}$$

If $b \equiv 0$ we will denote the differential operator by L_0 instead of L. The coefficients $a(x)$ and $b(x)$ are assumed to be measureable and to satisfy

$$0 < \alpha \le a(x) \le \beta < \infty, \qquad 0 < x < 1$$

$$0 \le b(x) \le \beta, \qquad 0 < x < 1.$$

Let $I = I(t_0,t_1) = \{x : t_0 < x < t_1\}$, where $0 \le t_0 < t_1 \le 1$. $I(0,1)$ will often be denoted by $\bar{I}$. By $W_p^k(I)$ with $k = 0,1,\ldots$ and $1 \le p \le \infty$ we denote the usual Sobolov spaces. $\overset{\circ}{W}{}_p^k I)$ denotes the subspace of $W_p^k(I)$ consisting of functions which vanish together with their derivatives of orders less than k at t_0 and t_1. On $W_p^k(I)$ we have the usual norms and semi-norms

$$\|u\|_{k,p,I} = \begin{cases} \left(\sum_{j=0}^{k} \|u^{(j)}\|_{L_p(I)}^p\right)^{1/p}, & 1 \le p < \infty \\ \max\limits_{0\le j\le k} \|u^{(j)}\|_{L_\infty(I)}, & p = \infty \end{cases}$$

and

$$|u|_{k,p,I} = \|u^{(k)}\|_{L_p(I)}.$$

We will also use the negatively indexed space $W_p^{-1}(I) = (\overset{\circ}{W}{}_q^1(I))'$ with the norm

$$\|u\|_{-1,p,I} = \sup_{v \in \overset{\circ}{W}{}_q^1(I)} \frac{|\int_I uv\,dx|}{\|v\|_{1,q,I}}$$

where $\frac{1}{p} + \frac{1}{q} = 1$. We will mostly work with $p = \infty$ and $p = 2$. If $p = 2$ we write

$$H^k(I) = W_2^k(I),$$

$$\|u\|_{k,I} = \|u\|_{k,2,I},$$

and

$$|u|_{k,I} = |u|_{k,2,I}.$$

For $r \geq 1$, $P^r(I)$ will denote the space of polynomials of degree less than or equal to r on I.

The solution of (2.1) will be understood in the usual weak sense in $\overset{\circ}{H}{}^1(\bar{I})$.

In our analysis we will use various a priori estimates for the Dirichlet problem. We now state these easily proved estimates.

Lemma 2.1. There is a constant $0 < C(\alpha,\beta) < \infty$ such that

(2.2) $$\|u\|_{1,\bar{I}} \leq C(\alpha,\beta)\|Lu\|_{0,\bar{I}}$$

for all u for which $Lu \in L_2(\bar{I})$ and $u(0) = u(1) = 0$, and

(2.3) $$\|u\|_{1,\infty,\bar{I}} \leq C(\alpha,\beta)\|Lu\|_{0,\infty,I}$$

for all u for which $Lu \in L_\infty(\bar{I})$ and $u(0) = u(1) = 0$.

Lemma 2.2. There are constants $0 < C(\alpha,\beta) < \infty$ and $0 < \tilde{C}(\alpha,\beta) < \infty$ such that

$$\tilde{C}(\alpha,\beta)\|Lu\|_{-1,\bar{I}} \leq \|u\|_{1,\bar{I}} \leq C(\alpha,\beta)\|Lu\|_{-1,\bar{I}} \tag{2.4}$$

for all $u \in \mathring{H}^1(\bar{I})$, and

$$\tilde{C}(\alpha,\beta)\|Lu\|_{-1,\infty,\bar{I}} \leq \|u\|_{1,\infty,\bar{I}} \leq C(\alpha,\beta)\|Lu\|_{-1,\infty,\bar{I}} \tag{2.5}$$

for all $u \in \mathring{W}^1_\infty(\bar{I})$.

3. Generalized Finite Element Methods

We will now consider a class of finite element methods for the solution of (2.1). Let $\Delta = \{0 = x_0 < x_1 < \ldots < x_n = 1\}$, where $n = n(\Delta)$ is a positive integer, be an arbitrary mesh on $\bar{I}$ and set $h_j = x_j - x_{j-1}$ and $I_j = I(x_{j-1}, x_j)$ for $j = 1,\ldots,n$, and $h = h(\Delta) = \max_j h_j$. With (2.1) we associate the bilinear forms

$$B_0(u,v) = \int_0^1 au'v'dx$$

and

$$B_1(u,v) = \int_0^1 buvdx$$

and the linear functional

$$F(v) = \int_0^1 fvdx.$$

The exact solution u of (2.1) is then characterized as the unique $u \in \mathring{H}^1(\bar{I})$ satisfying

$$B(u,v) \equiv B_0(u,v) + B_1(u,v) = F(v), \quad \text{for all} \quad v \in \mathring{H}^1(\bar{I}).$$

Given a finite dimensional space $S \subset \mathring{H}^1(\bar{I})$ we define an approximation $u_S \in S$ to u by

$$B_0(u_S,v) + B_1(u_S,v) = F(v), \quad \text{for all} \quad v \in S. \tag{3.1}$$

The usual choice for S is

$$S = S^r_{\Delta,0} = \{\psi \in \mathring{H}^1(\bar{I}) : \psi|_{I_j} \in P^r(I_j), \quad j = 1,\ldots,n\}$$

where $r = 1,2,\ldots$. Then $u_S = u_\Delta$ is the standard finite element approximation to u.

Remark. Sometimes the condition $S^r_{\Delta,0} \subset \overset{\circ}{H}{}^1(\bar{I})$ is relaxed and the bilinear forms B_0 and B_1 are defined as sums of integrals over the subintervals I_j. In this paper we will always assume $S^r_{\Delta,0} \subset \overset{\circ}{H}{}^1(\bar{I})$. However, the extension of the ideas developed in this paper to two dimensional boundary value problems will utilize such a generalization. This will be addressed in a forthcoming paper.

The finite element method based on $S^r_{\Delta,0}$ uses local basis functions, namely those functions in $S^r_{\Delta,0}$ that are characterized by nodal values together with $r - 1$ values associated with each I_j. We can in fact write

$$S^r_{\Delta,0} = S^1_{\Delta,0} \oplus \hat{S}^r_{\Delta,0}$$

where

$$\hat{S}^r_{\Delta,0} = \{\psi \in S^r_{\Delta,0} : \psi(x_j) = 0,\ j = 0,\ldots,n\}.$$

It is clear that $\dim S^1_{\Delta,0} = n - 1$ and $\dim \hat{S}^r_{\Delta} = (r-1)n$. $\hat{S}^r_{\Delta,0}$ is called the space of internal modes or a bubble space and is characterized in various ways, e.g., by function values at $r - 1$ interior points of each I_j or by values of the first $r - 1$ moments on each I_j. This is especially important when r is large. In this case hiearchical elements are of importance; for more on these elements see [5], e.g. Local basis functions lead to a "collection" process for the construction of the sparse global stiffness matrices from the local stiffness matrices (which are associated with the individual I_j). The load vector is constructed in a similar way. We see immediately that this essential feature of the standard finite element method is preserved if we replace the finite element equation (3.5) by

(3.2) $B_0(A_1u_\Delta, \mathcal{B}_1 v) + B_1(A_2 u_\Delta, \mathcal{B}_2 v) = F(\mathcal{B}_3 v)$, for all $v \in S_{\Delta,0}$,

where A_i, $i = 1,2$, and $\mathcal{B}_i$, $i = 1,2,3$, are one-to-one linear mappings of $S^r_{\Delta,0}$ into $\mathring{H}^1(\bar{I})$ with the property that for $\psi \in S^r_{\Delta,0}$, $A_i\psi|_{I_j}$ and $\mathcal{B}_i\psi|_{I_j}$ depends only on $\psi|_{I_j}$. For simplicity, we will in addition assume that

(3.3) $$(A_i\psi)(x_j) = \psi(x_j), \quad i = 1,2$$

and

(3.4) $$(\mathcal{B}_i\psi)(x_j) = \psi(x_j), \quad i = 1,2,3, \quad j = 1,\ldots,n.$$

(3.3) can be weakened while (3.4) is essential. Such a weakening is important in the two dimensional case. (Cf. the previous Remark.) We note that the standard finite element method corresponds to the choice $A_i = \mathcal{B}_i = E$ = the identity mapping. We will refer to the method defined by (3.6) as a <u>generalized finite element method</u> (GFEM). The method will be denoted by the vector $L = (A_1, A_2, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3)$ of mappings, and the corresponding finite element approximation by $u_\Delta(L)$. $e_\Delta(L) = u - u_\Delta(L)$ will denote the error.

<u>Remark</u>. The existence and uniqueness of $u_\Delta(L)$ as defined by (3.2) cannot be established without further assumptions. This question will be settled for each of the specific generalized finite element methods discussed in the paper.

Practically all finite element methods, including methods for solving diffusion-convection problems and reduced integration methods, can be put in the form (3.2). The form (3.2) obviously preserves the possibilities of the collection process for the construction of the stiffness matrix and preserves the sparsity of these matrices, so that one of the main ingredients of

the FEM is unchanged (see Section 1).

A number of questions arise in connection with generalized finite element methods:

i) What is the optimal choice for the mappings A_i, B_i, i.e., what is the best choice for the local stiffness matrices?

ii) How do implementational requirements influence the choice? How much effort should be spent in the construction of the A_i, B_i as compared with effort spent in mesh refinement?

iii) Which choices lead to robustness of the method?

iv) What is the accuracy of the generalized finite element method?

These questions will be addressed in the following sections.

4. The L_1-Finite Element Method

In this section we introduce a version of the GFEM which, although it is rather impractical, gives valuable insight into the best that could be expected from a method for solving (2.1) when the coefficients are rough.

In addition to the usual finite element space $S^r_{\Delta,0}$ introduced earlier we will use the spaces

$$\tilde{S}^r_\Delta \equiv \{\psi \in \overset{\circ}{H}{}^1(I) : L\psi|_{I_j} \in P^{r-2}(I_j) \quad \text{for each} \quad j\}$$

and $\tilde{S}^r_{\Delta,0} = \tilde{S}^r_\Delta \cap \overset{\circ}{H}{}^1(\bar{I})$. (If $r = 1$ we require that $L\psi = 0$ on each I_j.) We see that $\dim S^r_{\Delta,0} = \dim \tilde{S}^r_\Delta = nr - 1$. We now discuss several properties of the spaces $\tilde{S}^r_\Delta$.

Lemma 4.1. Given a function $u \in H^1(\bar{I})$ there is a unique $I^r_\Delta(a,b) = I_\Delta u$ satisfying

$$\text{(4.1)} \qquad \begin{cases} I_\Delta u \in \tilde{S}^r_\Delta \\ I_\Delta u(x_j) = u(x_j), \ j = 0,1,\ldots,n \\ \displaystyle\int_{I_j} (u - I_\Delta u)(x - x_{j-1})^\ell dx = 0, \ \ell = 0,\ldots,r-2, j = 1,\ldots,n. \end{cases}$$

$I_\Delta u \in \overset{\circ}{H}{}^1(\bar{I})$ if $u \in \overset{\circ}{H}{}^1(\bar{I})$.

Proof. The existence and uniqueness of $I_\Delta u$ will be established if we show that $u = 0$ in (4.1) implies $I_\Delta u = 0$. Thus suppose $u = 0$. Then from (4.1) and the definition of $\tilde{S}^r_\Delta$ we have

$$0 = \int_{I_j} I_\Delta u \; L(I_\Delta u)dx$$

$$= \int_{I_j} [a((I_\Delta u)')^2 + b(I_\Delta u)^2]dx,$$

from which we get $I_\Delta u = 0$.

We will refer to $I_\Delta u$ as the $\tilde{S}^r_\Delta$-interpolant of u.

<u>Lemma 4.2</u>. For $u \in \overset{\circ}{H}{}^1(\bar{I})$, $I_\Delta u \in \tilde{S}^r_{\Delta,0}$ is characterized by

(4.2) $$\int_0^1 (I_\Delta u)'v'dx = \int_0^1 u'v'dx, \quad \text{for all} \quad v \in S^r_{\Delta,0}$$

or by

(4.3) $$B(I_\Delta u,v) = B(u,v), \quad \text{for all} \quad v \in \tilde{S}^r_{\Delta,0}.$$

<u>Proof</u>. The characterizations (4.2) and (4.3) can be seen by writing

$$\int_0^1 (u-I_\Delta u)'v'dx = \textstyle\sum_j \int_{I_j} (u-I_\Delta u)(-v")dx + \sum_j (u-I_\Delta u)(x_j)Jv'(x_j)$$

and

$$B(u-I_\Delta u,v) = \textstyle\sum_j \int_{I_j} (u-I_\Delta u)Lvdx + \sum_j (u-I_\Delta u)(x_j)Jav'(x_j),$$

where $Jav'(x_j) = av'(x_j^-) - av'(x_j^+)$. We note that $Jav'(x_j)$ is well defined since $Lv|_{I_\ell} \in P^{r-2}(I_\ell)$ and therefore $av'|_{I_\ell} \in H^1(I_\ell)$ for $\ell = 1,\ldots,n$, which implies av' has left and right limits at each node x_j.

<u>Remark</u>. (4.3) shows that $I_\Delta u$ is the Ritz projection of $u \in \overset{\circ}{H}{}^1(\bar{I})$ onto the subspace $\tilde{S}^r_{\Delta,0}$ with respect to the form B.

<u>Lemma 4.3</u>. For $\ell = 0$ or 1, $k > 0$, and $r \geq 1$, we have

$$\|u - I_\Delta u\|_{\ell,\bar{I}} \le C(\alpha,\beta) h^{\mu} \|Lu\|_{k,\bar{I}} \tag{4.4}$$

and

$$\|u - I_\Delta u\|_{\ell,\infty,\bar{I}} \le C(\alpha,\beta) h^{\mu} \|Lu\|_{k,\infty,\bar{I}}\,, \tag{4.5}$$

where $\mu = \min(k+2-\ell, r+1-\ell)$, and $C(\alpha,\beta)$ is independent of u but depends in general on α, β, ℓ, k, and r.

Proof. We first prove (4.4) with $\ell = 0$. Suppose

$$\begin{cases} \tilde{L}w = -(\tilde{a}w')' + \tilde{b}w = F \text{ on } \bar{I} \\ w(0) = w(1) = 0 \end{cases}$$

where $\alpha \le \tilde{a}(x) \le \beta$, $0 \le \tilde{b}(x) \le \beta$ on $\bar{I}$, and let v be the Ritz projection of w onto the subspace

$$\tilde{S}_0 = \{\phi : \tilde{L}\phi \in P^{r-2}(\bar{I}),\ \phi(0) = \phi(1) = 0\}$$

with respect to the form $\tilde{B}(u,v) = \int_0^1 (\tilde{a}u'v'+\tilde{b}uv)dx$, i.e., suppose

$$\begin{cases} v \in \tilde{S}_0 \\ \tilde{B}(v,\phi) = \tilde{B}(w,\phi), \text{ for } \phi \in \tilde{S}_0\,. \end{cases}$$

We write $\tilde{a}$, $\tilde{b}$ for general coefficients in order to distinguish from a, b in (2.1) and $\bar{a}$, $\bar{b}$ introduced below. Now, using (2.2), (2.4), and the fact that all norms on the finite dimensional space $P^{r-2}(\bar{I})$ are equivalent for $2 \le p$ we have

$$(4.6)\qquad \|\tilde{L}v\|_{0,p,\bar{I}} \le C\|\tilde{L}v\|_{-1,\bar{I}} \le C(\alpha,\beta)\|v\|_{1,\bar{I}} \le C(\alpha,\beta)\|w\|_{1,\bar{I}} \le C(\alpha,\beta)\|F\|_{0,\bar{I}} \le C(\alpha,\beta)\|F\|_{0,p,\bar{I}} .$$

Now consider $u - I_\Delta u$ on I_j. For any function $v(x)$, $x \in I_j$, let $\bar{v}(\bar{x}) = v(x)$, where $\bar{x} = \frac{x-x_{j-1}}{h_j} \in \bar{I}$. By a standard scaling argument

$$(4.7)\qquad \|u - I_\Delta u\|_{0,I_j} \le Ch_j^{1/2}\|\bar{u} - \bar{I}\,\bar{u}\|_{0,\bar{I}}$$

where $\bar{I}\,\bar{u}$ is the $\tilde{S}^r$-interpolant of $\bar{u}$ where

$$\tilde{S}^r = \{\phi : \bar{L}\phi = -(\bar{a}\phi')' + h_j^2\bar{b}\phi \in P^{r-2}(\bar{I})\},$$

i.e., $\bar{I}\,\bar{u}$ is the element of $\tilde{S}^r$ having the same values as $\bar{u}$ at 0 and 1 and having the same moments of orders up to $r - 2$ as $\bar{u}$ (cf. (4.1)).

From (2.2) we see that

$$(4.8)\qquad \|\bar{u} - \bar{I}\,\bar{u}\|_{0,\bar{I}} \le C(\alpha,\beta)\|\bar{L}(\bar{u}-\bar{I}\,\bar{u})\|_{0,\bar{I}} .$$

Write

$$\bar{I}\,\bar{u} = (\bar{I}\,\bar{u})_1 + (\bar{I}\bar{u})_2$$

where

$$\begin{cases} \bar{L}(\bar{I}\,\bar{u})_1 = 0 \quad \text{on } \bar{I} \\ (\bar{I}\,\bar{u})_1(0) = \bar{u}(0), \ (\bar{I}\,\bar{u})_1(1) = \bar{u}(1). \end{cases}$$

It is easily seen that

$$\begin{cases} \bar{L}[\bar{u} - (I\bar{u})_1] = \bar{L}\bar{u} \quad \text{on} \quad \bar{I} \\ [\bar{u} - (\bar{I}\,\bar{u})_1]\,(0) = [\bar{u} - (\bar{I}\,\bar{u})_1]\,(1) = 0 \end{cases}$$

and that $(\bar{I}\,\bar{u})_2$ is the Ritz projection of $\bar{u} - (\bar{I}\,\bar{u})_1$ onto $\tilde{S}_0$ with respect to the form $\bar{B}(u,v) = \int_0^1 (\bar{a}u'v'+\bar{b}uv)dx$, where

$$\tilde{S}_0^r = \{\phi : \bar{L}\phi \in P^{r-2}(\bar{I}),\quad \phi(0) = \phi(1) = 0\}.$$

Hence from (4.6) with $w = \bar{u} - (\bar{I}\,\bar{u})_1$, $v = (\bar{I}\,\bar{u})_2$, and $p = 2$ we obtain

$$\text{(4.9)} \qquad \begin{aligned} \|\bar{L}(\bar{I}\,\bar{u})\|_{0,\bar{I}} &= \|\bar{L}(\bar{I}\,\bar{u})_2\|_{0,\bar{I}} \\ &\le C(\alpha,\beta)\|\bar{L}\bar{u}\|_{0,\bar{I}} \,. \end{aligned}$$

Combining (4.7), (4.8), and (4.9) we have

$$\text{(4.10)} \qquad \begin{aligned} \|u - I_\Delta u\|_{0,I_j} &\le C(\alpha,\beta)h_j^{1/2}\|\bar{L}\bar{u}\|_{0,\bar{I}} \\ &\le C(\alpha,\beta)h_j^{1/2}\|\bar{L}\bar{u}\|_{\mu-2,\bar{I}} \end{aligned}$$

since $\mu \ge 2$ for $\ell = 0$. Now

$$\|u - I_\Delta u\|_{0,I_j} = \|(u-\phi) - I_\Delta(u-\phi)\|_{0,I_j}$$

for any ϕ such that $L\phi|_{I_j} \in P^{r-2}(I_j)$. Combining this with (4.10) we see that

$$\|u - I_\Delta u\|_{0,I_j} \le C(\alpha,\beta)h_j^{1/2}\|\bar{L}\bar{u} - \bar{L}\bar{\phi}\|_{\mu-2,\bar{I}} \,.$$

Since $\bar{L}\bar{\phi}$ is arbitrary in $P^{r-2}(\bar{I})$ we have

$$\|u - I_\Delta u\|_{0,I_j} \le C(\alpha,\beta) h_j^{1/2} \inf_{Q \in P^{r-2}} \|\bar{L}\bar{u} - Q\|_{\mu-2,\bar{I}}$$

$$\le C(\alpha,\beta) h_j^{1/2} |\bar{L}\bar{u}|_{\mu-2,\bar{I}} .$$

Then a further scaling argument yields

$$\|u - I_\Delta u\|_{0,I_j} \le C(\alpha,\beta) h^\mu |Lu|_{\mu-2,I_j} . \tag{4.11}$$

(4.4) with $\ell = 0$ follows immediately from (4.11).

(4.4) with $\ell = 1$ is obtained by a slight modification of the above argument. We finally note that (4.5) follows from a similar argument based on (2.3) and (4.6) with $p = \infty$.

Lemma 4.4. There is a constant $C(\alpha,\beta)$, independent of u and Δ, such that

$$\|I_\Delta u\|_{1,\bar{I}} \le C(\alpha,\beta) \|u\|_{1,\bar{I}} \tag{4.12}$$

and

$$\|I_\Delta u\|_{1,\infty,\bar{I}} \le C(\alpha,\beta) \|u\|_{1,\infty,\bar{I}} . \tag{4.13}$$

Proof. (4.12) follows directly from the coercivity of B. We turn to the proof of (4.13).

Let $z \in W^1_\infty(\bar{I})$ and suppose w satisfies

$$\begin{cases} \tilde{L}w = -(\tilde{a}w')' + \tilde{b}w = Q \in P^{r-2}(\bar{I}) \text{ on } \bar{I} \\ w(0) = z(0),\ w(1) = z(1) \\ \int_0^1 wx^j dx = \int_0^1 zx^j dx, \quad j = 0,\dots,r-2 \end{cases}$$

where $0 < \alpha \le \tilde{a}(x) \le \beta$, $0 \le \tilde{b}(x) \le \beta$. Write $w = w_1 + w_2$ where

$$\begin{cases} \tilde{L}w_1 = 0 \\ w_1(0) = z(0), \quad w_1(1) = z(1) \end{cases}$$

and

$$\begin{cases} \tilde{L}w_2 = Q \\ w_2(0) = w_2(1) = 0. \end{cases}$$

We also write $z = z_1 + z_2$ where

$$\begin{cases} z_1 \in H^1(\bar{I}), \; z_1(0) = z(0), \; z_1(1) = z(1) \\ \tilde{B}(z_1,\phi) = 0, \quad \text{for} \quad \phi \in \mathring{H}^1(\bar{I}) \end{cases}$$

and

$$\begin{cases} z_2 \in \mathring{H}^1(\bar{I}) \\ \tilde{B}(z_2,\phi) = \tilde{B}(z,\phi), \quad \text{for} \quad \phi \in \mathring{H}^1(\bar{I}), \end{cases}$$

where $\tilde{B}(u,v) = \int_0^1(\tilde{a}u'v'+\tilde{b}uv)dx$. $\tilde{B}(z_1,z_2) = 0$ and hence

(4.14) $$\|z_2\|_{1,\bar{I}} \leq C(\alpha,\beta)\|z\|_{1,\bar{I}} .$$

Observing that w_2 is the Ritz projection of z_2 onto the subspace

$$\{\phi : \tilde{L}\phi = P^{r-2}(\bar{I})\phi , \quad \phi(0) = \phi(1) = 0\}$$

we see that

(4.15) $$\|w_2\|_{1,\bar{I}} \leq C(\alpha,\beta)\|z_2\|_{1,\bar{I}} .$$

We now show that

(4.16) $$\|w\|_{1,\infty,\bar{I}} \leq C(\alpha,\beta)\|z\|_{1,\infty,\bar{I}} .$$

First we show that

$$\|w_1\|_{1,\infty,\bar{I}} \le C(\alpha,\beta)\|z\|_{0,\infty,\bar{I}} . \tag{4.17}$$

If $\tilde{b} = 0$,

$$w_1(x) = \frac{z(1)-z(0)}{\int_0^1 \frac{dt}{\tilde{a}}} \int_0^x \frac{dt}{\tilde{a}} + z(0)$$

and the result is immediate. Now suppose $\tilde{b} \neq 0$. Let v satisfy

$$\begin{cases} \tilde{L}_0 v = -(\tilde{a}v')' = 0 \\ v(0) = z(0),\ v(1) = z(1). \end{cases}$$

From (4.17) with $\tilde{b} = 0$ we have

$$\|v\|_{1,\infty,\bar{I}} \le C(\alpha,\beta)\|z\|_{0,\infty,\bar{I}} . \tag{4.18}$$

Letting $y = w_1 - v$ we see that

$$\begin{cases} \tilde{L}y = -\tilde{b}v \\ y(0) = y(1) = 0. \end{cases}$$

Using (2.3) and (4.18) we obtain

$$\|y\|_{1,\infty,\bar{I}} \le C(\alpha,\beta)\|v\|_{0,\infty,\bar{I}} .$$

(4.17) follows immediately from this result and (4.18).

Next we show that

$$\|w_2\|_{1,\infty,\bar{I}} \le C(\alpha,\beta)\|z\|_{1,\infty,\bar{I}} . \tag{4.19}$$

Using (2.3), (2.4), and the equivalence of norms on the finite dimensional space $P^{r-2}(\bar{I})$ we have

$$(4.20)\qquad \|w_2\|_{1,\infty,\bar{I}} \le C(\alpha,\beta)\|Q\|_{0,\infty,\bar{I}} \le C(\alpha,\beta)\|Q\|_{-1,\bar{I}} \le C(\alpha,\beta)\|w_2\|_{1,\bar{I}}\,.$$

Combining (4.14), (4.15), and (4.20) we obtain (4.19). (4.16) follows from (4.17) and (4.19).

Now consider $u - I_\Delta u$ on I_j and recall the notation introduced in the proof of Lemma 4.3. By a scaling argument we have

$$(4.21)\qquad |u - I_\Delta u|_{1,\infty,I_j} \le Ch_j^{-1}|\bar{u} - \bar{I}\,\bar{u}|_{1,\infty,\bar{I}}$$

where $\bar{I}\,\bar{u}$ is the $\tilde{S}^r$-interpolant of $\bar{u}$ where

$$\tilde{S}^r = \{\phi : \bar{L}\phi = -(\bar{a}\,\phi')' + h_j^2\bar{b}\phi \in P^{r-2}(\bar{I})\}.$$

From (4.16) and (4.21) we obtain

$$(4.22)\qquad |u - I_\Delta u|_{1,\infty,I_j} \le C(\alpha,\beta)h_j^{-1}\|\bar{u}\|_{1,\infty,\bar{I}}\,.$$

From (4.22) and (4.5) we have

$$(4.23)\qquad |u-I_\Delta u|_{1,\infty,I_j} \le |(u-\gamma)+I_\Delta(u-\gamma)|_{1,\infty,I_j}+|\gamma||1-I_\Delta 1|_{1,\infty,I_j} \le C(\alpha,\beta)[h_j^{-1}\|\bar{u}-\gamma\|_{1,\infty,\bar{I}} + |\gamma|h_j]$$

for any constant γ. Now select $\gamma = \int_0^1 \bar{u}\,dx$. Then

$$\|\bar{u} - \gamma\|_{1,\infty,\bar{I}} \le C|\bar{u}|_{1,\infty,\bar{I}}$$

and

$$|\gamma| \le |\bar{u}|_{0,\infty,\bar{I}}\,.$$

Thus from (4.23) we get

$$(4.24)\qquad |u - I_\Delta u|_{1,\infty,I_j} \le C(\alpha,\beta)[h_j^{-1}|\bar{u}|_{1,\infty,\bar{I}} + h_j|\bar{u}|_{0,\infty,\bar{I}}].$$

A second scaling argument applied to (4.24) yields

$$\begin{aligned}|u - I_\Delta u|_{1,\infty,I_j} &\le C(\alpha,\beta)[|u|_{1,\infty,I_j} + h_j|u|_{0,\infty,I_j}] \\ &\le C(\alpha,\beta)\|u\|_{1,\infty,I_j}\end{aligned}$$

and thus

$$(4.25)\qquad |I_\Delta u|_{1,\infty,I_j} \le C(\alpha,\beta)\|u\|_{1,\infty,I_j}.$$

(4.13) follows immediately from (4.25).

The main goal of this section is to consider the L_1-method, where $L_1 = (A,A,A,A,A)$, the mapping $A : S^r_{\Delta,0} \to \mathring{H}^1(\bar{I})$ being defined by $Au = I^r_\Delta(a,b)u$ for $u \in S^r_{\Delta,0}$. It is immediate that A is one-to-one and onto $\tilde{S}^r_{\Delta,0}$, $Au|_{I_j}$ depends only on $u|_{I_j}$. $Au(x_j) = u(x_j)$, and that $A^{-1}w = I^r_\Delta(1,0)w$ for $w \in \tilde{S}^r_{\Delta,0}$. The existence and uniqueness of $u_\Delta(L_1)$ is clear. We see that $u \in \hat{S}^r_{\Delta,0}$ implies $Au(x_j) = 0$ and that

$$AS^r_{\Delta,0} = AS^1_{\Delta,0} \oplus A\hat{S}^r_{\Delta,0}.$$

This fact plays an important role in implementation.

<u>Lemma 4.5</u>. There are constants C and $C(\alpha,\beta)$, independent of Δ, such that

$$(4.26)\qquad \|Au\|_{1,\bar{I}} \le C(\alpha,\beta)\|u\|_{1,\bar{I}},$$

and

$$(4.27)\qquad \|Au\|_{1,\infty,\bar{I}} \le C(\alpha,\beta)\|u\|_{1,\infty,\bar{I}}, \quad \text{for } u \in S^r_{\Delta,0},$$

and

(4.28) $$\|A^{-1}w\|_{1,\bar{I}} \le C\|w\|_{1,\bar{I}},$$

and

(4.29) $$\|A^{-1}w\|_{1,\infty,\bar{I}} \le C\|w\|_{1,\infty,\bar{I}}, \quad \text{for} \quad w \in S^r_{\Delta,0}.$$

Proof. (4.26) and (4.27) are essentially restatements of (4.12) and (4.13), respectively. (4.28) and (4.29) are special cases of (4.26) and (4.27), respectively, corresponding to $a = 1$ and $b = 0$.

We present the error estimates for the L_1-method in the next two theorems. Let $\tilde{e}_\Delta(L_1) = u - Au_\Delta(L_1)$.

Theorem 4.1. For $\ell = 0,1$ we have

(4.30) $$\|\tilde{e}_\Delta(L_1)\|_{\ell,\bar{I}} \le C(\alpha,\beta)h^\mu\|f\|_{k,\bar{I}}$$

and

(4.31) $$\|e_\Delta(L_1)\|_{\ell,\infty,\bar{I}} \le C(\alpha,\beta)h^\mu\|f\|_{k,\infty,\bar{I}}$$

where $\mu = \min(k+2-\ell,\ r+1-\ell)$.

Proof. It follows immediately from the definition of the L_1-method and from (4.3) that $Au_\Delta(L_1) = I_\Delta u$. Thus (4.30) and (4.31) follow directly from Lemma 4.3.

Theorem 4.2. For $\ell = 0,1$ we have

(4.32) $$\|e_\Delta(L_1)\|_{\ell,\bar{I}} \leq C(\alpha,\beta)h^{1-\ell}\|f\|_{0,\bar{I}}$$

and

(4.33) $$\|e_\Delta(L_1)\|_{\ell,\infty,\bar{I}} \leq C(\alpha,\beta)h^{1-\ell}\|f\|_{0,\infty,\bar{I}}.$$

Proof. These results follow immediately from Theorem 4.1, the fact that $u_\Delta(L_1)$ is the $S^r_{\Delta,0}$-interpolant of $Au_\Delta(L_1)$, a standard approximation result for $S^r_{\Delta,0}$, and (4.26) and (4.27).

Remarks.

1) The rates of convergence given in (4.32) and (4.33) are clearly the highest that could be proved under the governing hypotheses.

2) Suppose that we have found $u_\Delta(L_1)$. Then, although the accuracy of $u_\Delta(L_1)$ is low (see Theorem 4.2 and Remark 1), $Au_\Delta(L_1)$ has high accuracy (see Theorem 4.1) and can be calculated locally, element by element. Thus the highly accurate $Au_\Delta(L_1)$ can be constructed by post processing.

3) Theorem 4.1 shows that we can obtain the same rate of convergence in the norms $\|\cdot\|_{\ell,\bar{I}}$, $\ell = 0,1$, for problems with rough coefficients as for problems with smooth coefficients when post processing is applied. This shows the maximum possible robustness.

4) It is easy to show that postprocessing does not improve the rate of convergence if a and b are smooth and so $e_\Delta(L_1)$

and $e_\Delta(L_1)$ are of the same order in this case.

5) The L_1-method is closely related to Ritz method using L-splines (see [6], e.g.) and, as would thus be expected, yields the exact solution at the nodes.

6) We have presented estimates in only the norms $\|\cdot\|_{\ell,\infty,\bar{I}}$ and $\|\cdot\|_{\ell,\bar{I}}$ for $\ell = 0$ and 1. Parallel results hold in the norms $\|\cdot\|_{\ell,p,\bar{I}}$, $2 \le p \le \infty$, $\ell = 0,1$.

5. The L_0-Finite Element Method

In this section we consider the L_0-method, where $L_0 = (A_0, A_0, A_0, A_0, A_0)$, the mapping A_0 being constructed from τ_0 as A was from L in Section 4, i.e., $A_0 u = I_\Delta^r(a,0)u$ for $u \in S^r_{\Delta,0}$. The existence and uniqueness of $u_\Delta(L_0)$ is immediate.

Theorem 5.1. For $\ell = 0$ or 1 we have

$$(5.1) \qquad \|u - A_0 u_\Delta(L_0)\|_{\ell,\bar I} \le C(\alpha,\beta) h^{2-\ell} \|f\|_{k,\bar I}$$

and

$$(5.2) \qquad \|u - A_0 u_\Delta(L_0)\|_{\ell,\infty,\bar I} \le C(\alpha,\beta) h^{2-\ell} \|f\|_{k,\infty,\bar I}.$$

Proof. Let $F = f - bu$. Then from (2.2) and (2.3)

$$(5.3) \qquad \|F\|_{0,\bar I} \le C(\alpha,\beta) \|f\|_{0,\bar I}$$

and

$$(5.4) \qquad \|F\|_{0,\infty,\bar I} \le C(\alpha,\beta) \|f\|_{0,\infty,\bar I}.$$

Clearly

$$(5.5) \qquad \begin{cases} L_0 u = F \\ u(0) = u(1) = 0. \end{cases}$$

Let $\bar u_\Delta(L_0)$ denote the L_0-finite element approximation to the solution of (5.5). From Theorem 4.1 applied to (5.5), (5.3) and (5.4) we have

$$(5.6) \qquad \|u - A_0 \bar u_\Delta(L_0)\|_{\ell,\bar I} \le C(\alpha,\beta) h^{2-\ell} \|f\|_{0,\bar I}$$

and

$$\|u-A_0\bar{u}_\Delta(L_0)\|_{\ell,\infty,\bar{I}} \le C(\alpha,\beta)h^{2-\ell}\|f\|_{0,\infty,\bar{I}}. \tag{5.7}$$

Writing

$$u_\Delta(L_0) = \bar{u}_\Delta(L_0) + z(L_0) \tag{5.8}$$

we immediately see that

$$\begin{aligned} B(A_0z,A_0v) &= B(A_0u_\Delta(L_0),A_0v) \\ &\quad - B_0(A_0\bar{u}_\Delta(L_0),A_0v) - B_1(A_0\bar{u}_\Delta(L_0),A_0v) \\ &= B_1(u-A_0\bar{u}_\Delta(L_0),A_0v), \quad \text{for} \quad v \in S^r_{\Delta,0}. \end{aligned} \tag{5.9}$$

From (5.9) and the coercivity of B we have

$$\begin{aligned} \|A_0z\|^2_{1,\bar{I}} &\le C(\alpha,\beta)B(A_0z,A_0z) \\ &\le C(\alpha,\beta)\|u-A_0\bar{u}_\Delta(L_0)\|_{0,\bar{I}}\|A_0z\|_{0,\bar{I}} \end{aligned}$$

and this, together with (5.6), yields

$$\|A_0z\|_{1,\bar{I}} \le C(\alpha,\beta)h^2\|f\|_{0,\bar{I}}. \tag{5.10}$$

(5.1) follows from (5.6), (5.8), and (5.10).

For $\ell = 0$, (5.2) follows from (5.7), (5.10) and the fact that

$$\|A_0z\|_{0,\infty,\bar{I}} \le \|A_0z\|_{1,\bar{I}}. \tag{5.11}$$

It remains to prove (5.2) for $\ell = 1$. It is easily shown that there is a positive constant $C(\alpha,\beta)$ such that

$$(5.12) \qquad \sup_{u\in \overset{\circ}{W}{}^1_\infty(\bar{I})} \frac{|B_0(u,v)|}{\|u\|_{1,\infty,\bar{I}}} \geq C(\alpha,\beta)\|v\|_{1,1,\bar{I}}, \quad \text{for} \quad u \in \overset{\circ}{W}{}^1_1(\bar{I})$$

From (4.3), (4.13), and (5.12) we have

$$(5.13) \qquad \sup_{w\in A_0 S^r_{\Delta,0}} \frac{|B_0(w,v)|}{\|w\|_{1,\infty,\bar{I}}} = \sup_{u\in \overset{\circ}{W}_\infty(\bar{I})} \frac{|B_0(I_\Delta(a,0)u,v)|}{\|I_\Delta(a,0)u\|_{1,\infty,\bar{I}}}$$

$$= \sup_{u\in \overset{\circ}{W}{}'_\infty(\bar{I})} \frac{|B_0(u,v)|}{\|u\|_{1,\infty,\bar{I}}} \quad \frac{\|u\|_{1,\infty,\bar{I}}}{\|I_\Delta(a,0)u\|_{1,\infty,\bar{I}}}$$

$$\geq C(\alpha,\beta)\|v\|_{1,1,\bar{I}}, \quad \text{for} \quad v \in A_0 S^r_\Delta .$$

(5.13), together with the fact that an operator and its adjoint have equal norms, yields

$$(5.14) \qquad \sup_{v\in A_0 S^r_{\Delta,0}} \frac{|B_0(w,v)|}{\|v\|_{1,1,\bar{I}}} \geq C(\alpha,\beta)\|w\|_{1,\infty,\bar{I}}, \quad \text{for} \quad w \in A_0 S^r_\Delta .$$

Now from (5.7), (5.9), (5.10),(5.11), and (5.14) we obtain

$$(5.15) \qquad \|A_0 z\|_{1,\infty,\bar{I}} \leq C(\alpha,\beta) \sup_{v\in S^r_{\Delta,0}} \frac{|B_0(A_0 z, A_0 v)|}{\|A_0 v\|_{1,1,\bar{I}}}$$

$$= C(\alpha,\beta) \sup_{v\in S^r_{\Delta,0}} \frac{|B_1(u - A_0\bar{u}\ (L_0) - A_0 z, A_0 v)|}{\|A_0 v\|_{1,1,\bar{I}}}$$

$$\leq C(\alpha,\beta)(\|u - A_0\bar{u}_\Delta(L_0)\|_{0,\infty,\bar{I}} + \|A_0 z\|_{0,\infty,\bar{I}})$$

$$\leq C(\alpha,\beta)h^2\|f\|_{0,\infty,\bar{I}} .$$

(5.2) with $\ell = 1$ follows directly from (5.7), (5.8), and (5.15).

Remarks.

1. The rate $2-\ell$ in (5.1) and (5.2) cannot be improved. This can be seen by examining the special case $a = 1$ and b rough.

2. (5.2) with $\ell = 0$ shows that

$$|u(x_j)-u_\Delta(L_0)(x_j)| \le C(\alpha,\beta)h^2\|f\|_{0,\infty,\bar{I}}, \quad j = 1,\ldots,n-1.$$

We also have the analogue of Theorem 4.2.

Theorem 5.2. For $\ell = 0$ or 1 we have

$$\|e_\Delta(L_0)\|_{\ell,\bar{I}} \le C(\alpha,\beta)h^{1-\ell}\|f\|_{0,\bar{I}} \tag{5.16}$$

and

$$\|e_\Delta(L_0)\|_{\ell,\infty,\bar{I}} \le C(\alpha,\beta)h^{1-\ell}\|f\|_{0,\infty,\bar{I}}. \tag{5.17}$$

Remark. Comparing Theorems 5.1 and 5.2 we see that the L_0 and L_1 methods have the same rates of convergence. On the other hand, implementation of the L_0 method is much simpler since the equation $L_0\psi = \phi$ can be solved by quadrature.

6. The L_2-Finite Element Method

In this section we consider the $L_2 \equiv (A_0, E, E, E, E)$ method, where E is the identity mapping and A_0 is as defined in Section 5. First we analyze the case $b = 0$.

Lemma 6.1. Suppose $b = 0$. Then for $\ell = 0,1$ we have

(6.1) $$\|u - A_0 u_\Delta(L_2)\|_{\ell,\bar{I}} \leq C(\alpha,\beta) h^\mu \|f\|_{k,\bar{I}}$$

and

(6.2) $$\|u - A_0 u_\Delta(L_2)\|_{\ell,\infty,\bar{I}} \leq C(\alpha,\beta) h^\mu \|f\|_{k,\infty,\bar{I}}$$

where $\mu = \min(k+1, r)$.

Proof. We will prove (6.1) in detail for $r = 2$. In $S^2_{\Delta,0}$ we construct a basis by choosing

$$\phi_1, \ldots, \phi_{n-1}$$

$$\xi_1, \ldots, \xi_n$$

satisfying

$$\phi_i(x_j) = \delta_{ij}, \quad i,j = 1,\ldots,n-1,$$

$$\int_{I_j} \phi_i dx = 0, \quad i = 1,\ldots,n-1, \; j = 1,\ldots,n,$$

$$\xi_i(x_j) = 0, \quad i = 1,\ldots,n, \; j = 1,\ldots,n-1,$$

$$\int_{I_j} \xi_i dx = \delta_{ij}, \quad i,j = 1,\ldots,n.$$

For $f \in H^k(\bar{I})$ let

$$f_\Delta = \sum_{j=1}^{n-1} \left(\int_{\bar{I}} f\, \phi_j\, dx\right) \delta_{x_j} + \sum_{j=1}^{n} \left(\int_{\bar{I}} f\, \xi_j\, dx\right) \chi_j$$

where δ_{x_j} is the Dirac distribution at x_j and χ_j is the characteristic function of I_j. It is easily shown that $A_0 u_\Delta(L_2)$ is the exact solution of problem (2.1) with right hand side f_Δ. We note that this fact implies the existence and uniqueness of $u_\Delta(L_2)$. Thus from (2.4) we have

$$\|u - A_0 u_\Delta(L_2)\|_{1,\bar{I}} \le C(\alpha,\beta)\|f - f_\Delta\|_{-1,\bar{I}}. \tag{6.3}$$

Now let w be the solution of

$$\begin{cases} -w'' = f, & x \in \bar{I} \\ w(0) = w(1) = 0 \end{cases} \tag{6.4}$$

and let w_Δ be the standard finite element approximation to w in the subspace $S^2_{\Delta,0}$ i.e., let w_Δ be the Ritz projection of w onto $S^2_{\Delta,0}$ with respect to the form $\int_0^1 u'v'dx$. As above, we see that w_Δ is the exact solution of (6.4) with right hand side f_Δ. Thus from (2.4), applied to the problem (6.4), we obtain

$$\|f - f_\Delta\|_{-1,\bar{I}} \le C\|w - w_\Delta\|_{1,\bar{I}}. \tag{6.5}$$

Combining (6.3), (6.5), a well known estimate for the standard finite element method, and a simple regularity estimate for (6.4) we have

$$\|u-A_0u_\Delta(L_2)\|_{1,\bar{I}} \le C(\alpha,\beta)h^2\|w\|_{k+2,\bar{I}}$$
$$\le C(\alpha,\beta)h^2\|f\|_{k,\bar{I}}.$$

This is (6.1) for $r = 2$. The proof for general r is similar. Finally we note that the proof of (6.2) follows the same lines but is based on (2.5) instead of (2.4).

Remark. In general one cannot obtain a better rate of convergence in $H^0(\bar{I})$ $(W^0_\infty(\bar{I}))$ than one obtains in $H^1(\bar{I})$ $(W^1_\infty(\bar{I}))$.

Lemma 6.2. Suppose $b = 0$. Then for $\ell = 0$ or 1 we have

(6.6) $$\|e_\Delta(L_2)\|_{\ell,\bar{I}} \le C(\alpha,\beta)h^{1-\ell}\|f\|_{0,\bar{I}}$$

and

(6.7) $$\|e_\Delta(L_2)\|_{\ell,\bar{I}} \le C(\alpha,\beta)h^{1-\ell}\|f\|_{0,\infty,\bar{I}}.$$

Proof. The proof of this lemma is similar to the proof of Theorem 4.2.

We now turn to the general case.

Theorem 6.1. There is a constant $h_0 = h_0(\alpha,\beta)$ such that for $0 < h < h_0$, $u_\Delta(L_0)$ is uniquely determined. Furthermore there is a constant $C(\alpha,\beta)$ such that

(6.8) $$\|e_\Delta(L_2)\|_{0,\bar{I}} \le C(\alpha,\beta)h\|f\|_{0,\bar{I}}$$

and

(6.9) $$\|e_\Delta(L_2)\|_{0,\infty,I} \le C(\alpha,\beta)h\|f\|_{0,\infty,\bar{I}}$$

for $0 < h < h_0$.

Proof. Consider the problem

$$\begin{cases} -(aw')' = F, & x \in \bar{I} \\ w(0) = w(1) = 0. \end{cases}$$

Let T be the corresponding solution operator, i.e., let $TF = w$, and let T_Δ be the L_2 approximate solution operator, i.e., let $T_\Delta F = w_\Delta(L_2)$. We regard T and T_h as operators on $L_2(\bar{I})$. From Lemma 6.2 we have

$$\|T - T_\Delta\|_{L_2,L_2} \le C(\alpha,\beta)h. \tag{6.10}$$

Let u be the exact solution of (2.1) and $u_\Delta(L_2)$ the L_2 approximation to u. Then

$$u = T(f - bu) \tag{6.11}$$

and

$$u_\Delta(L_2) = T_\Delta(f - b\, u_\Delta(L_2)). \tag{6.12}$$

Using (6.11) and (6.12) we obtain

$$\begin{aligned} u - u_\Delta(L_2) &= (T-T_\Delta)f - Tb(u-u_\Delta(L_2)) \\ &\quad + (T_\Delta-T)b(u_\Delta(L_2)-u) + (T_\Delta-T)bu \end{aligned} \tag{6.13}$$

which can be written as

$$[I + Tb + (T_\Delta-T)b]e_\Delta(L_2) = (T-T_\Delta)\ (f - bu) \tag{6.14}$$

where I is the identity operator on $L_2(\bar{I})$. Now TbI is compact on L_2 and it is easily seen that $I + TbI$ is one-to-one. Thus $(I + TbI)^{-1}$ exists and is bounded. From (6.10) we see that

$$\|(T_\Delta - T)bI\|_{L_2,L_2} \leq C(\alpha,\beta)h$$

and hence

$$(6.15) \qquad \|[I + TbI + (T_\Delta - T)bI]^{-1}\|_{L_2,L_2} \leq C(\alpha,\beta)$$

for h sufficiently small. Identity (6.14) together with the invertibility of $[I + TbI + (T_\Delta - T)bI]$ implies the existence and uniqueness of $u_\Delta(L_2)$. It follows immediately from (6.10), (6.14), and (6.15) that

$$\|e_\Delta(L_2)\|_{0,\bar{I}} \leq C(\alpha,\beta)h\|f - bu\|_{0,\bar{I}}$$
$$\leq C(\alpha,\beta)h\|f\|_{0,\bar{I}}.$$

This is (6.8). The proof of (6.9) is similar.

Remark.

1. The assumption that h is small is not necessary as will be shown in Section 8. See Remark 2 following Theorem 8.2.

2. In Lemma 6.2 the case $\ell = 1$ is essentially a stability result, this result is also valid in the general case (when $b \not\equiv 0$). See Remark 2 following Theorem 8.2.

Theorem 6.2. For $\ell = 0$ or 1 we have

(6.18) $$\|u-A_0u_\Delta(L_2)\|_{\ell,\bar{I}} \le C(\alpha,\beta)h\|f\|_{0,\bar{I}}$$

and

(6.19) $$\|u-A_0u_\Delta(L_2)\|_{\ell,\infty,\bar{I}} \le C(\alpha,\beta)h\|f\|_{0,\infty,\bar{I}}.$$

Proof. Recall the notations from the proof of Theorem 6.1. Now let $\tilde{T}_\Delta$ be defined by $\tilde{T}_\Delta F = A_0 T_\Delta F$. Just as we obtained (6.13) we now obtain

(6.20) $$\begin{aligned} u - A_0u_\Delta(L_2) &= (T-\tilde{T}_\Delta)f - Tb(u-u_\Delta(L_2)) \\ &\quad + (\tilde{T}_\Delta-T)b(u_\Delta(L_2)-u) \\ &\quad + (\tilde{T}_\Delta-T)bu. \end{aligned}$$

From Lemma 6.1 we have

(6.21) $$\|T-\tilde{T}_\Delta\|_{L_2,H^1} \le C(\alpha,\beta)h\|f\|_{0,\bar{I}}.$$

Using (2.2) we see that

(6.22) $$\|Tb(u-u_\Delta(L_2))\|_{1,\bar{I}} \le C(\alpha,\beta)\|u-u_\Delta(L_2)\|_{0,\bar{I}}.$$

Finally, combining (6.20)-(6.22) and Theorem 6.1 we see that

$$\|u - A_0u\ (L_2)\|_{1,\bar{I}} \le C(\alpha,\beta)h\|f\|_{0,\bar{I}}.$$

This gives (6.18). The proof of (6.19) is similar.

Remarks.

1. Estimates (6.6) and (6.8) cannot be improved when general a and b are present. If a and b are smooth then, of course, better estimates can be derived, namely

$$(6.23) \qquad \|e_\Delta(L_2)\|_{\ell,\bar{I}} \le C\|f\|_{k,\bar{I}}\, h^\mu .$$

where $\ell = 0$ or 1 and $\mu = \min(k+2-\ell,\ r+1-\ell)$. Here C depends on the smoothness of a and b as well as on α and β. Similar results hold in connection with the norm $\|\cdot\|_{\ell,\infty,\bar{I}}$.

2. (6.16) and (6.17) cannot in general be improved. Theorem 6.2 shows, however, that the accuracy can be improved by post processing.

3. The rate of convergence in (6.23) is the best possible, as can be seen from the theory of n-widths (see [7]). If a and b are merely measurable it is also possible to show that (6.8) and (6.16) give the best possible rate of convergence. Thus we see that $u_\Delta(L_2)$ yields the optimal rate of convergence in all cases. Similar results hold in connection with the norm $\|\cdot\|_{\ell,\infty,\bar{I}}$.

4. The L_2 method is well suited for implementation. In fact we will see in Section 8 that it can be obtained as a mixed method or can be easily implemented directly.

5. We could also discuss other choices for the A_i and B_i which would modify in various ways the accuracy of the approximate solution.

7. The Standard Finite Element Method (The L_3-method)

The standard finite element (or Ritz) method corresponds to the choice $L_3 = (E,E,E,E,E)$. In this section we discuss a negative result that shows the contrast with the L_2-method. We assume $r = 1$.

Consider the family of boundary value problems given by a_n, $b = 0$, $f = 1$, where $\Delta_n = \{x_0,\ldots,x_n\} = \{0,\frac{1}{n},\ldots,\frac{n-1}{n},1\}$ and

$$a_n(x) = \begin{cases} 2, & x_{j-1} < x < \frac{x_{j-1}+x_j}{2} \\ 1, & \frac{x_{j-1}+x_j}{2} < x < x_j, \quad j = 1,\ldots,n \end{cases}$$

and let u_n denote the exact solution. It is easy to see that $u_{\Delta_n}(L_3)$ (the standard finite element approximation to u_n) is the piecewise linear interpolant to the exact solution of the problem

$$\begin{cases} -\frac{3}{2}v'' = 1 \\ v(0) = v(1) = 0 \,. \end{cases}$$

From Lemma 6.2 we have

$$\|u_n - u_{\Delta_n}(L_2)\|_{0,\bar{I}} \leq \frac{C}{n}$$

and it can be shown (see Theorem 9.3) that u_n converges in L_2 to the exact solution of

$$\begin{cases} -\frac{4}{3} w'' = 1 \\ w(0) = w(1) = 0. \end{cases}$$

This shows that the L_2 approximation, $u_{\Delta_n}(L_2)$, is very accurate while the standard finite element approximation, $u_{\Delta_n}(L_3)$, fails to "converge", i.e., $u_n - u_{\Delta_n}(L_2) \not\to 0$.

Remarks.

1. We see that the standard method is very nonrobust. It does not converge for rough coefficients at all, and gives very poor results when a changes significantly between the nodel points. In contrast the L_2 method performs well whether the coefficients are smooth or rough. Because the implementation of the L_2 involves the same amount of work as does the standard method, the L_2-method should be prefered in all cases.

2. We have analyzed only linear boundary value problems. Many of the ideas introduced here are valid for nonlinear problems also. For strongly nonlinear problems the L_2 method is very promising.

8. The Relation Between Generalized Finite Element Methods and Mixed Methods

In this section we discuss the relation between Generalized Finite Element Methods and mixed methods. Mixed methods for the approximate solution of (2.1) can be derived from the displacement formulation (2.1) by introducing an auxiliary variable, writing the second order scalar equation as a first order system, casting the systems into variational form, and then discretizing the resulting variational equations. This can be done in various ways. We consider two mixed methods for (2.1).

Letting $s = pu'$ we write (2.1) as a first order system in two ways, namely as

$$\begin{cases} au' - s = 0, & 0 < x < 1 \\ -s' + bu = f, & 0 < x < 1 \\ u(0) = u(1) = 0 & \end{cases}$$

and

$$\begin{cases} u' - \dfrac{s}{a} = 0, & 0 < x < 1 \\ -s' + bu = f, & 0 < x < 1 \\ u(0) = u(1) = 0, & \end{cases}$$

and then consider the associated variational formulations

$$(8.1)\quad \begin{cases} u \in \mathring{H}^1(\bar{I}), \quad \sigma \in H^\circ(\bar{I}) \\ \int_0^1 au'\sigma dx - \int_0^1 s\sigma dx = 0, \quad \text{for all} \quad \sigma \in H^\circ(\bar{I}) \\ \int_0^1 sv'dx + \int_0^1 buvdx = \int_0^1 fvdx, \quad \text{for all} \quad v \in \mathring{H}^1(\bar{I}) \end{cases}$$

and

$$(8.2)\quad \begin{cases} u \in \mathring{H}^1(\bar{I}), \quad \in H^\circ(\bar{I}) \\ \int_0^1 u'\sigma dx - \int_0^1 \frac{s\sigma}{a} dx = 0, \quad \text{for all} \quad \sigma \in H^\circ(\bar{I}) \\ \int_0^1 sv'dx + \int_0^1 buvdx = \int_0^1 fvdx, \quad \text{for all} \quad v \in \mathring{H}^1(\bar{I}). \end{cases}$$

The mixed methods we consider are now obtained by discretizing (8.1) and (8.2):

$$(8.3)\quad \begin{cases} u_\Delta \in S^r_{\Delta,0}, s_\Delta \in \hat{S}^{r-1}_\Delta \\ \int_0^1 au'_\Delta\sigma dx - \int_0^1 s_\Delta\sigma dx = 0, \quad \text{for all} \quad \sigma \in \hat{S}^{r-1}_\Delta \\ \int_0^1 s_\Delta v'dx + \int_0^1 bu_\Delta vdx = \int_0^1 fvdx, \quad \text{for all} \quad v \in S^r_{\Delta,0}, \end{cases}$$

$$(8.4)\quad \begin{cases} u_\Delta \in S^r_{\Delta,0}, s_\Delta \in \hat{S}^{r-1}_\Delta \\ \int_0^1 u'_\Delta\sigma dx - \int_0^1 \frac{s\sigma}{a} dx, \quad \text{for all} \quad \sigma \in \hat{S}^{r-1}_\Delta \\ \int_0^1 s_\Delta v'dx + \int_0^1 bu_\Delta vdx = \int_0^1 fvdx, \quad \text{for all} \quad v \in S^r_{\Delta,0}, \end{cases}$$

where $\hat{S}_\Delta^{r-1} = \{\psi \in H^\circ(I): \psi/_{I_j} = P^{r-1}(I_j),\ j = 1,\dots,n\}$.

As indicated above, the mixed formulations (8.1) and (8.2) are obtained from the displacement formulation (2.1) via the introduction of an auxiliary variable. Clearly this process can be reversed, i.e., (2.1) can be obtained from (8.1) or (8.2) by eliminating the variable s. It is also possible to eliminate this variable at the discrete level, i.e., to eliminate s_Δ from (8.3) or (8.4), obtaining a Generalized Finite Element Method. This is made precise in the next two theorems.

Theorem 8.1. The elimination of s_Δ from (8.3) leads to the L_3 finite element method for u_Δ, i.e., u_Δ is the standard finite element approximation to u.

Proof. Suppose u_Δ, s_Δ satisfies (8.3). Letting $\sigma = v'$ in the first equation in (8.3) we find $\int_0^1 a u_\Delta' v' dx = \int_0^1 s_\Delta v' dx$. Using this in the second equation we see that

$$\begin{cases} u_\Delta \in S_{\Delta,0}^r \\ \displaystyle\int_0^1 a u_\Delta' v' dx + \int_0^1 b u_\Delta v dx = \int_0^1 f v dx, \quad \text{for all } v \in S_{\Delta,0}^r, \end{cases}$$

i.e., u_Δ is the L_3 finite element approximation to u.

Theorem 8.2. The elimination of s_Δ from (8.4) leads to the L_2 finite element method for u_Δ.

Proof. Suppose u_Δ, s_Δ satisfies (8.4). From the first equation in (8.4) we see that

$$u'_\Delta\big|_{I_j} = P\left(\frac{s_\Delta}{a}\bigg|_{I_j}\right)$$

where P is the L_2-projection onto $P^{r-1}(I_j)$. The mapping $s_\Delta\big|_{I_j} \to R_j(s_\Delta\big|_{I_j}) = P\left(\frac{s_\Delta}{a}\bigg|_{I_j}\right)$ from $P^{r-1}_{(I_j)}$ to $P^{r-1}_{(I_j)}$ is one-to-one; let $Q_j = R_j^{-1}$. Then

$$s_\Delta\big|_{I_j} = Q_j u'_\Delta\big|_{I_j}.$$

Using this in the second equation in (8.4) we see that u_Δ satisfies

$$\begin{cases} u_\Delta \in S^r_{\Delta,0} \\ \sum_{j=1}^n \int_{I_j} Q_j u'_\Delta\, v'dx + \int_0^1 bu_\Delta v dx = \int_0^1 f v dx, \quad \text{for all } v \in S^r_{\Delta,0}. \end{cases}$$

It remains to show that $Q_j u'_\Delta\big|_{I_j} = a(A_0u_\Delta)'\big|_{I_j}$. From (4.2) we see that $P(A_0\varphi)'\big|_{I_j} = \varphi'\big|_{I_j}$ for $\varphi \in S^r_{\Delta,0}$ and from the definition of $\tilde{S}^r_\Delta$ we have $(A_0u)'\big|_{I_j} \in P^{r-1}(I_j)$. Thus

$$\begin{aligned} R_j[a(A_0u_\Delta)'\big|_{I_j}] &= P(A_0u_\Delta)'\big|_{I_j} \\ &= u'_\Delta\big|_{I_j}. \end{aligned}$$

This completes the proof.

Remarks.

1. If $r = 1$ both the L_2 and L_3 methods can be viewed as arising from the application of the standard finite element method to a problem with appropriately perturbed coefficients. Consider the problem.

$$\begin{cases} -(a_\Delta u')' + bu = f, & 0 \in \bar{I} \\ u(0) = u(1) = 0 \end{cases}$$

where a_Δ is constant on each I_j, and consider its approximation by the standard finite element method. If $a_\Delta = \bar{a}_\Delta$, where

$\bar{a}_\Delta\Big|_{I_j} = \dfrac{\int_{I_j} a\,dx}{h_j}$, i.e., if a_Δ is the piecewise average of a,

we obtain the standard finite element method for the solution of (2.1). Also, it is clear (when $r = 1$) that

$$Q_j u'_\Delta\Big|_{I_j} = \frac{u'_\Delta\big|_{I_j} h_j}{\int_{I_j} \frac{dx}{a}}$$

where Q_j is the operator introduced in the proof of Theorem 8.2.

Hence if $a_\Delta = \bar{\bar{a}}_\Delta$, where $\bar{\bar{a}}_\Delta\Big|_{I_j} = \left(\dfrac{\int_{I_j} \frac{dx}{a}}{h_j}\right)^{-1}$, i.e., if a_Δ is the piecewise harmonic average of a, we get the L_2 method. It is obvious that $\bar{a}_\Delta$ and $\bar{\bar{a}}_\Delta$ can differ significantly if a changes significantly on the subintervals I_j, while for smooth a, $\bar{a}_\Delta$ and $\bar{\bar{a}}_\Delta$ differ by $O(h^2)$.

The observation that the L_2 method is tha same as the

standard method applied to the problem with a replaced by $\bar{\bar{a}}_\Delta$, where $\bar{\bar{a}}_\Delta$ is the piecewise harmonic average of a, shows that the L_2 method can be easily implemented, its implementation being of the same difficulty as for the standard method.

2) Using the operator R_j introduced in the proof of Theorem 8.2 we can prove the existence and uniqueness of $u_\Delta(L_2)$ for all h. See Remark 1 following Theorem 6.1. R_j is positive definite on $P^{r-1}(I_j)$ with respect to the L_2-inner product and its eigenvalues are bounded away from ∞ and 0 uniformly in j. Thus

$$\int_{I_j} (Q_j z) z dx \geq \lambda \int_{I_j} z^2 dx, \quad \text{for} \quad z \in P^{r-1}(I_j)$$

with $\lambda > 0$ independent of j. Now suppose

$$\int_0^1 a[A_0 u_\Delta(L_2)]'v' + \int b u_\Delta(L_2) v \, dx = 0, \quad \text{for all} \quad v \in S^r_{\Delta,0}.$$

Then

$$0 = \sum_j \int_{I_j} [Q_j u'_\Delta(L_2)] u'_\Delta(L_2) \, dx + \int_0^1 b[u_\Delta(L_2)]^2$$

$$\geq \lambda \sum_j \int_{I_j} [u'_\Delta(L_2)]^2 dx$$

which implies $u_\Delta(L_2) = 0$.

3) Nemat-Nasser, in [8] and a subsequent series of papers, suggested the use of the variational principle (8.2) in conjunction with trigonometric polynomial approximating functions for the approximate calculation of eigenvalues of problems with

rough coefficients. The authors derived error estimates for this method and corresponding finite element methods (cf. (8.4)) in [9].

9. Additional Properties of the L_2-method

So far in this paper we have assumed the coefficients $a(x)$ and $b(x)$ are in $L_\infty(\bar{I})$. In this section we derive additional properties of the L_2-method in the case in which $a(x)$ has bounded variation and $r = 1$. We begin by proving a perturbation of coefficient result.

Let u_a denote the solution of (2.1), i.e., let u_a be the solution of

$$(9.1) \qquad \begin{cases} -(au_a')' + bu_a = f, x \in \bar{I} \\ u_a(0) = u_a(1) = 0 \end{cases}$$

and let u_{a_Δ} be the solution of the corresponding problem with a replaced by a_Δ, where a_Δ is the piecewise harmonic average of a subordinate to the mesh $\Delta = \{0 = x_1 < x_2 < \cdots < x_n = 1\}$, i.e., the step function defined by

$$a_\Delta\Big|_{I_j} = \left(\frac{\int_{I_j} \frac{dx}{a}}{h_j} \right)^{-1} .$$

We now present an estimate for $u_a - u_{a_\Delta}$ under the hypothesis that $a(x)$ has bounded variation.

For a function $c(x)$ with bounded variations denote by c^Δ the piecewise average of c. We will assume functions with bounded variation to be right continuous at each $x < 1$ and left continuous at 1. It is easily seen that $V(c^\Delta) \le V(c)$, where $V = V_{\bar{I}}$ denotes the variation on $\bar{I}$.

<u>Lemma 9.1</u>. There is a constant C such that

$$(9.2) \qquad \left| \int_0^{x_j} (c-c^\Delta)\phi dx \right| \le CV_{\bar{I}}(c)|\phi|_{1,p,\bar{I}}\, h^{2-\frac{1}{p}}$$

for $\phi \in W_p^1(\bar{I})$, $1 \le p \le \infty$, $j = 1,\ldots,n$.

Proof. Let $\phi \in W_p^1(\bar{I})$, let ϕ^Δ be the piecewise average of ϕ, and set $\Phi(x) = \int_0^x (\phi-\phi^\Delta)dt$. Then

$$
\begin{aligned}
(9.3) \qquad \left| \int_0^{x_j} (c-c^\Delta)\phi dx \right| &= \left| \int_0^{x_j} (c-c^\Delta)(\phi-\phi^\Delta)dx \right| \\
&= \left| \int_0^{x_j} (c-c^\Delta)\Phi'(x)dx \right| \\
&= \left| \int_0^{x_j} (c-c^\Delta)d\Phi \right| \\
&= \left| \int_0^{x_j} \Phi d(c-c^\Delta) \right| \\
&\le C\|\Phi\|_{0,\infty,\bar{I}}\ V(c-c^\Delta) \\
&\le C\|\Phi\|_{0,\infty,\bar{I}}\ V(c).
\end{aligned}
$$

Now we bound $\|\Phi\|_{0,\infty,\bar{I}}$. Suppose $x_{\ell-1} \le x < x_\ell$. Then

$$|\Phi(x)| = \left| \int_0^{x_{\ell-1}} (\phi-\phi^\Delta)\ dt + \int_{x_{\ell-1}}^{x} (\phi-\phi^\Delta)dt \right|$$

$$= \left| \int_{x_{\ell-1}}^{x} (\phi-\phi^{\Delta})\,dt \right|$$

$$\leq \|\phi-\phi^{\Delta}\|_{0,p,I_\ell}(x-x_{\ell-1})^{\frac{p-1}{p}}$$

and thus by the Poincaré inequality we have

$$(9.4) \qquad |\phi(x)| \leq h_\ell \|\phi'\|_{0,p,I_\ell} h_\ell^{\frac{p-1}{p}}$$

$$= \|\phi'\|_{0,p,I_\ell} h^{2-\frac{1}{p}}.$$

(9.2) follows from (9.3) and (9.4).

For $b \equiv 0$ we obtain two estimates for $u_a - u_{a_\Delta}$. The first is an estimate for $\|u_a - a_{u_\Delta}\|_{0,p,\bar{I}}$ and the second is an estimate for $u_a(x_j) - u_{a_\Delta}(x_j)$, $j = 1,\ldots,n-1$.

Theorem 9.2. Suppose $b \equiv 0$. There is a constant $C(\alpha,\beta)$ depending only on α and β satisfying

$$(9.5) \qquad \left| \int_{\bar{I}} (u_a - u_{a_\Delta}) z\,dx \right| \leq C(\alpha,\beta) V_{\bar{I}}(a) \|z\|_{0,q,\bar{I}} \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}}$$

for $z \in L_p(\bar{I})$, $1 \leq p \leq \infty$, where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. The solution u_a of (9.1) is given by

$$(9.6) \qquad u_a(x) = -D_{a,f}(x) + \frac{D_{a,f}(1) E_a(x)}{E_a(1)}$$

where

$$D_{a,f}(x) = \int_0^x \frac{\int_0^t f(s)\,ds}{a(t)}\,dt$$

and

$$E(x) = \int_0^x \frac{dt}{a(t)},$$

and the solution u_{a_Δ} is given by

$$(9.7) \qquad u_{a_\Delta}(x) = -D_{a_\Delta,f}(x) + \frac{D_{a_\Delta,f}(1)E_{a_\Delta}(x)}{E_{a_\Delta}(1)}.$$

We first estimate $\int_0^1 (D_{a,f}-D_{a_\Delta,f})z\,dx$.

Set $Q(x) = \int_1^x z(t)dt$. Then

$$\int_0^1 (D_{a,f}-D_{a_\Delta,f})z\,dx = \int_0^1\left(\int_0^x \left(\frac{1}{a(t)} - \frac{1}{a_\Delta(t)}\right)\int_0^t f(s)ds\,dt\right)z(x)\,dx$$

$$= -\int_0^1 \left(\frac{1}{a(x)} - \frac{1}{a_\Delta(x)}\right)\int_0^x f(s)ds\,Q(x)\,dx.$$

Noting that $\frac{1}{a_\Delta} = (\frac{1}{a})^\Delta$ we can apply Lemma 9.1 with $c = \frac{1}{a}$, $\phi(x) = \int_0^x f\,ds\,Q(x)$, and $x_j = 1$ to get

$$(9.8) \quad \int_0^1 (D_{a,f}-D_{a_\Delta,f})z\,dx \le CV(\tfrac{1}{a})\,|\phi|_{1,q,\bar{I}}\,h^{1+\frac{1}{p}}\,\|z\|_{0,q,\bar{I}}$$

$$\le C(\alpha)\ V(a)\,\|f\|_{0,q,\bar{I}}\,\|z\|_{0,q,\bar{I}}\,h^{1+\frac{1}{p}}.$$

Next we estimate $\int_0^1 (E-E_\Delta)z\,dx$. We see that

$$\int_0^1 (E_a - E_{a_\Delta})z\,dx = \int_0^1 \int_0^x \left(\frac{1}{a} - \frac{1}{a_\Delta}\right)dt\, z(x)\,dx$$

$$= -\int_0^1 \left(\frac{1}{a(x)} - \frac{1}{a_\Delta(x)}\right) Q(x)\,dx.$$

Now we apply Lemma 9.1 with $c = \frac{1}{a}$ and $\phi(x) = Q(x)$ and obtain

$$(9.9) \qquad \int_0^1 (E_a - E_{a_\Delta})z\,dx \leq C(\alpha) V_{\bar{I}}(a) \|z\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}}.$$

Applying Lemma 9.1 with $c = \frac{1}{a}$ and $\phi(x) = \int_0^x f(s)ds$ we see that

$$(9.10) \qquad |D_{a,f}(1) - D_{a_\Delta,f}(1)| = \left| \int_0^1 \left(\frac{1}{a} - \frac{1}{a_\Delta}\right) \int_0^x f(s)ds\,dt \right|$$

$$\leq C(\alpha) V_{\bar{I}}(a) \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}}.$$

From (9.6) and (9.7) and the fact that $E_a(1) = E_{a_\Delta}(1)$ we see that

$$(9.11) \qquad u_a(x) - u_{a_\Delta}(x)$$

$$= D_{a_\Delta,f}(x) - D_{a,f}(x) + E_a(1)^{-1}[D_{a,f}(1)E_a(x) - D_{a_\Delta,f}(1)E_{a_\Delta}(x)]$$

$$= [D_{a_\Delta,f}(x) - D_{a,f}(x)] + E_a(1)^{-1}\{[D_{a,f}(1) - D_{a_\Delta,f}(1)]E_a(x)$$

$$+ D_{a_\Delta,f}(1)[E_a(x) - E_{a_\Delta}(x)]\}$$

(9.5) follows immediately from (9.11) and (9.8)-(9.10).

Theorem 9.2. Suppose $q \geq 0$. Then there is a constant $C(\alpha)$ such that

$$(9.12) \qquad |u_a(x_j) - u_{a_\Delta}(x_j)| \leq C(\alpha) V_{\bar{I}}(a) \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}} .$$

Proof. From (9.11) we see that

$$u_a(x_j) - u_{a_\Delta}(x_j)$$

$$= D_{a_\Delta,f}(x_j) - D_{a,f}(x_j) + E_a(1)^{-1}[D_{a,f}(1) - D_{a_\Delta,f}(1)]E_a(x_j).$$

Applying Lemma 9.1 with $c = \frac{1}{a}$ and $\phi(x) = \int_0^x f(s)ds$ we obtain (9.12).

We now consider the general case in which $b \not\equiv 0$.

Theorem 9.3. There is a constant $C(\alpha,\beta)$ such that

$$(9.13) \qquad \|u_a - u_{a_\Delta}\|_{0,p,\bar{I}} \leq C(\alpha,\beta) V_{\bar{I}}(a) \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}}$$

for $1 \leq p \leq \infty$.

Proof. If $b \equiv 0$ we can apply (9.5) to get

$$\|u_a - u_{a_\Delta}\|_{0,p,\bar{I}} = \sup_{z \in L_{\frac{p}{p-1}}(\bar{I})} \frac{|\int_0^1 (u_a - u_{a_\Delta}) z\, dx|}{\|z\|_{0,\frac{p}{p-1},\bar{I}}}$$

$$\leq C(\alpha) V_{\bar{I}}^{1}(a) \|f\|_{0,\frac{p}{p-1},\bar{I}}\, h^{1+\frac{1}{p}} .$$

If we let

$$L_0u \;=\; -(au')',$$

$$L_{\Delta,0}u \;=\; -(a_\Delta u')',$$

and

$$Bu \;=\; b(x)u$$

this result can be written as

$$(9.14) \qquad \|L_0^{-1} - L_{\Delta,0}^{-1}\|_{L_q(\bar{I}),L_p(\bar{I})} \;\le\; C(\alpha)\; V(a)h^{1+\frac{1}{p}},$$

where $\frac{1}{p} + \frac{1}{q} = 1$, and the desired result amounts to proving

$$\|(L_0+B)^{-1} - (L_{\Delta,0}+B)^{-1}\|_{L_q(\bar{I}),L_p(\bar{I})} \;\le\; C(\alpha)\; V(a)h^{1+\frac{1}{p}}.$$

We observe that

$$(9.15) \quad (L_\Delta+B)^{-1} - (L+B)^{-1}$$

$$= \quad (L_\Delta+B)^{-1}(L+B)(L+B)^{-1} - (L+B)^{-1}$$

$$= \quad (L_\Delta+B)^{-1}L_\Delta L_\Delta^{-1}[I-B(L+B)^{-1}] - [-(L_\Delta+B)^{-1}B+I]L^{-1}L(L+B)^{-1}$$

$$= \quad [I-(L_\Delta+B)^{-1}B]L_\Delta^{-1}[I-B(L+B)^{-1}] - [-(L_\Delta+B)^{-1}B+I]L^{-1}[I-B(L+B)^{-1}]$$

$$= \quad [I-(L_\Delta+B)^{-1}B](L_\Delta^{-1}-L^{-1})[I-B(L+B)^{-1}].$$

It is easily seen that

$$\|I - (L_\Delta + B)^{-1}B\|_{L_p(\bar{I}), L_p(\bar{I})} \leq C(\alpha,\beta) \tag{9.16}$$

and

$$\|I - B(L+B)^{-1}\|_{L_q(\bar{I}), L_q(\bar{I})} \leq C(\alpha,\beta). \tag{9.17}$$

It now follows immediately from (9.14)-(9.17) that

$$\begin{aligned} &\|u_a - u_{a_\Delta}\|_{0,p,\bar{I}} \\ &= \|(L_\Delta+B)^{-1}f - (L+B)^{-1}f\|_{0,p,\bar{I}} \\ &\leq \|I - (L_\Delta+B)^{-1}B\|_{L_p,L_p} \|L_\Delta^{-1} - L^{-1}\|_{L_q(\bar{I}),L_p(\bar{I})} \|I - B(L+B)^{-1}\|_{L_q,L_q} \\ &\qquad \cdot \|f\|_{0,q,\bar{I}} \\ &\leq C(\alpha,\beta) V_{\bar{I}}(a) \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}}. \end{aligned}$$

This completes the proof.

We now return to a consideration of the L_2-method under the assumption $r = 1$.

<u>Theorem 9.4</u>. There is a constant $C(\alpha,\beta)$ such that

$$\|u - u_\Delta\|_{0,p,\bar{I}} \leq C(\alpha,\beta) V_{\bar{I}}(p) \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}}. \tag{9.18}$$

<u>Proof</u>. Using Theorem 9.3 we have

(9.19)

$$\|u - u_\Delta\|_{0,p,\bar{I}} \le \|u_a - u_{a_\Delta}\|_{0,p,\bar{I}} + \|u_{a_\Delta} - u_\Delta\|_{0,p,\bar{I}}$$

$$\le C(\alpha,\beta) V_{\bar{I}}(a) h^{1+\frac{1}{p}} \|f\|_{0,q,\bar{I}} + \|u_{a_\Delta} - u_\Delta\|_{0,p,\bar{I}} \,.$$

We next observe that u_Δ can be viewed as the standard finite element approximation to u_{a_Δ} (cf. Remark 1 following Theorem 8.2). If $b \equiv 0$, u_Δ is the piecewise linear interpolant of u_{a_Δ} and by standard approximation results we have

$$\|u_{a_\Delta} - u_\Delta\|_{0,p,\bar{I}} \le C(\alpha,\beta) h^{1+\frac{2}{p}} \|f\|_{0,q,\bar{I}} \,. \tag{9.20}$$

If $b \not\equiv 0$, (9.20) can be proved by a slight modification of the usual duality argument. (9.18) follows directly from (9.19) and (9.20).

Remark. The estimate given in (9.18) is optimal.

Theorem 9.5. Suppose $b \equiv 0$. Then there is a constant C such that

$$|u(x_j) - u_\Delta(x_j)| \le C(\alpha) V_{\bar{I}}(a) \|f\|_{0,q,\bar{I}}\, h^{1+\frac{1}{p}} \tag{9.21}$$

for $1 \le p \le \infty$.

Proof. This result follows immediately from Theorem 9.2 and the fact that $u_\Delta(x_j) = u_{a_\Delta}(x_j)$.

<u>Remarks</u>.

1) Theorem 9.4 gives the rate of convergence of the L_2-method without post processing. It is possible to prove that the same result holds for the L_1-method.

2) One can prove that under the hypotheses of Theorem 9.4 we have the rate of convergence $h^{1/p}$ in the norm $\|\cdot\|_{1,p,\overline{I}}$.

10. Some Illustrious Computations

In this section we illustrate the improvement offered by the L_2-method over the L_3-method (the standard finite element method) by considering two examples.

Example 1. Consider problem (2.1) with

$$a(x) = \begin{cases} a_\ell , & 0 < x < 1/2 \\ a_r , & 1/2 < x < 1 \end{cases}$$

and

$$b(x) = 0$$

where a_ℓ and a_r are positive constants and consider the approximation of the solution u by the L_2-approximation $u_{\Delta_n}(L_2)$ and the L_3-approximation $u_{\Delta_n}(L_3)$ corresponding to the uniform mesh Δ_n with an odd number of subintervals $(x_j = jh,\ j = 0,1,\ldots,n,$ $h = n^{-1}$, n odd$)$. One can obtain exact expressions for the errors at the nodes. For the node just to the left of $\frac{1}{2}(x_{(n-1)/2})$ we find

$$e_{\Delta_n}(L_3)(\tfrac{1}{2}-\tfrac{h}{2}) = u(\tfrac{1}{2}-\tfrac{h}{2}) - u_{\Delta_n}(L_3)(\tfrac{1}{2}-\tfrac{h}{2})$$

$$= \frac{h(a_r-a_\ell)^3}{8a_\ell(a_\ell+a_r)^3} + \frac{h^2(a_\ell-a_r)}{8a_\ell(a_\ell+a_r)}\{8a_\ell a_r(a_\ell^2+a_r^2) + (a_\ell^2-a_r^2)\}$$

and

$$e_{\Delta_n}(L_2)(\tfrac{1}{2}-\tfrac{h}{2}) = u(\tfrac{1}{2}-\tfrac{h}{2}) - u_{\Delta_n}(L_2)(\tfrac{1}{2}-\tfrac{h}{2})$$

$$= \frac{h^2(a_\ell-a_r)}{8a_\ell(a_\ell+a_r)} + \frac{h^3(a_r-a_\ell)}{8a_\ell(a_\ell+a_r)} .$$

From these expressions we see that if $a_r - a_\ell$ is large relative to h, then $e_\Lambda(L_2) \ll e(L_3)$, while if $a_r - a_\ell$ is approximately equal to h, then $e_\Lambda(L_2) \sim e_\Lambda(L_3) \sim \frac{-h^3}{16a_\ell^2}$. Tables 1 and 2 give numerical values of the errors at the nodal points for the case $a_\ell = 1$, $a_r = 100$.

x	$e_{\Delta_n}(L_3)$	$e_{\Delta_n}(L_2)$
.2	9.23-3	-1.96-3
.4	1.85-2	-3.92-3
.6	-1.85-4	3.92-5
.8	-9.23-5	1.96-5

Table 1. $e_{\Delta_n}(L_3)$ and $e_{\Delta_n}(L_2)$ for $a_\ell = 1$, $a_r = 100$, $n = 5$.

x	$e_{\Delta_n}(L_3)$	$e_{\Delta_n}(L_2)$
1/21	5.32-4	-2.65-5
2/21	1.06-3	-5.29-5
3/21	1.60-3	-7.94-5
4/21	2.13-3	-1.06.4
5/21	2.66-3	-1.32-4
6/21	3.19-3	-1.59-4
7/21	3.72-3	-1.85-4
8/21	4.25-3	-2.12-4
9/21	4.79-3	-2.38-4
10/21	5.32-3	-2.65-4
11/21	-5.31-5	2.71-6
12/21	-4.78-5	2.44-6
13/21	-4.25-5	2.18-6
14/21	-3.71-5	1.92-6
15/21	-3.18-5	1.65-6
16/21	-2.65-5	1.38-6
17/21	-2.12-5	1.11-6
18/21	-1.59-5	8.35-7
19/21	-1.06-5	5.58-7
20/21	-5.30-6	2.84-7

Table 2. $e_{\Delta_n}(L_3)$ and $e_{\Delta_n}(L_2)$ for $a_\ell = 1$, $a_r = 100$, $n = 21$.

Example 2. Here we consider an example with a smooth but rapidly varying coefficient. Let

$$a(x) = 6 + \frac{500(x-\frac{1}{2})}{1+100|x-\frac{1}{2}|}, \qquad b(x) = 0,$$

and consider the same mesh as in Example 1. To give some idea of the variation of $a(x)$ we present a table of its values (Table 3).

x	a(x)
.0	1.10
.1	1.12
.2	1.16
.3	1.24
.4	1.45
.5	6.00
.6	10.55
.7	10.76
.8	10.84
.9	10.88
1.0	10.90

Table 3. Values of $a(x)$.

Table 4 gives the errors at the node just to the left of 1/2 for the standard method (L_3) and the L_2-method for this example.

n	$e_{\Delta_n}(L_3)$	$e_{\Delta_n}(L_2)$
9	1.48-3	-4.31-4
15	6.31-4	-1.15-4
21	3.34-4	-4.60-5
27	2.00-4	-2.28-5
33	1.30-4	-1.28-5
39	9.03-5	-7.75-6
45	6.53-5	-5.12-6

Table 4. $e_{\Delta_n}(L_3)$ and $e_{\Delta_n}(L_2)$ at the node just to the left of 1/2.

11. Conclusions

In this section we summarize in tabular form some of the results we have obtained on the convergence rates for the various methods and then make several remarks on conclusions that can be drawn. The convergence rates in the table are for smooth f.

	L_3-method (standard FE method)		L_1-method		L_2-method	
	$\|e\|_{H^0}$	$\|e\|_{H^1}$	$\|e\|_{H^0}$	$\|e\|_{H^1}$	$\|e\|_{H^0}$	$\|e\|_{H^1}$
a,b smooth	h^{r+1}	h^r	h^{r+1}	h^r	h^{r+1}	h^r
a,b measurable, without post-processing	h^0	h^0	h	h^0	h	h^0
a,b measurable, with post-processing	h^0	h^0	h^{r+1}	h^r	h	h
a bounded variation, b measurable, r=1	h^0	h^0	$h^{3/2}$	$h^{1/2}$	$h^{3/2}$	$h^{1/2}$

Remarks.

1) The standard FEM is very nonrobust.

2) The rates for the L_1-method show that by investing a considerably larger effort in the construction of the micro-stiffness matrices, one can significantly increase the accuracy--as compared with the standard method. This suggests that the goal of minimizing the entire computational effort with respect to desired accuracy could well be served by increasing the percentage of total effort that is spent at the micro level. This

could be especially important in convention with adaptive procedures.

3) The computational effort involved in implementing the L_2-method is exactly the same as with the standard method. Thus the L_2-method should clearly be preferred over the standard method in all cases.

4) The Generalized Finite Element Method offers a much larger freedom in the choice of computational procedures than does the standard finite element method, and as a consequence offers the possibility of significant improvement in accuracy, especially when used in conjunction with adaptive procedures.

REFERENCES

[1] Ch. F. Williamson, A history of the finite element method to the middle 1960's. Boston University Graduate School, Ph.D. 1976, History of Sciences.

[2] B. Fredrikson and J. Mackerle, Structural Mechanics Finite Element Computer Programs. A decade with Finite Elements 1970-1980. ABC Advanced Engineering Corp. Box 3044, S-580 D3 Linköping Sweden.

[3] I. Babuška and W. G. Szymczak, An Error Analysis for the Finite Element Method Applied to Convection-Diffusion Problems. To appear in Computer Methods in Appl. Math. Eng.

[4] J. E. Osborn, The numerical solution of differential equations with rough coefficients, in Advances in Computer Methods for Partial Differential Equations-IV. Proceedings of the Fourth IMAC International Symposium on Computer Methods in Partial Differential Equations 1981 (R. Vichnevetsky, R. S. Stepleman, Ed.), pp. 9-13.

[5] B. A. Szabo, Some recent developments in finite element analysis, Comp. Math. Appl. 5 (1979), pp. 99-115.

[6] R. S. Varga, Functional Analysis and Approximation Theory in Numerical Analysis, Regional Conference Series in Applied Mathematics, No. 3, SIAM, 1971.

[7] I. Babuška and A. K. Aziz, Survey lectures on the mathematical foundation of the finite element method, in The Mathematical Foundation of the Finite Element Method with Applications to Partial Differential Equations (A. K. Aziz, Ed.) Academic Press, New York, 1972, pp. 5-359

[8] S. Nemat-Nasser, Harmonic waves in layered composites. J of Appl. Mech., 39, Trans. ASME, vol. 94, Ser. E (1972), pp. 850-852.

[9] I. Babuska and J. E. Osborn, Numerical treatment of eigenvalue problems for differential equations with discontinuous coefficients, Math. Comput. 32 (1978), pp. 991-1023.

The Laboratory for Numerical Analysis is an integral part of the Institute for Physical Science and Technology of the University of Maryland, under the general administration of the Director, Institute for Physical Science and Technology. It has the following goals:

- To conduct research in the mathematical theory and computational implementation of numerical analysis and related topics, with emphasis on the numerical treatment of linear and nonlinear differential equations and problems in linear and nonlinear algebra.

- To help bridge gaps between computational directions in engineering, physics, etc. and those in the mathematical community.

- To provide a limited consulting service in all areas of numerical mathematics to the University as a whole, and also to government agencies and industries in the State of Maryland and the Washington Metropolitan area.

- To assist with the education of numerical analysts, especially at the postdoctoral level, in conjunction with the Interdisciplinary Applied Mathematics Program and the programs of the Mathematics and Computer Science Departments. This includes active collaboration with government agencies such as the National Bureau of Standards.

- To be an international center of study and research for foreign students in numerical mathematics who are supported by foreign governments or exchange agencies (Fulbright, etc.).

Further information may be obtained from Professor I. Babuška, Chairman, Laboratory for Numerical Analysis, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742.

FILMED

5-8